Optimizing Euclidean Distance Thresholds in Deep Facial Embeddings for High-Accuracy Real-Time Attendance Tracking

Prathamesh Sankpal[1], Atharva Mahangade[2], Nikhil Jakapure[3], Samarth Telkar[4]
[1,2,3,4]*Department of Electrical & Computer Engineering (ECE),*
Bharati Vidyapeeth Deemed to be University College of Engineering Pune, Maharashtra

Abstract— **The automated system for tracking presence through computer vision has become one of the most important innovations for the problems faced by many schools and offices. Although deep learning models produce very accurate facial embeddings, their use in the real world demands careful optimization of classification thresh-olds to guarantee high levels of both accuracy and convenience. This paper introduces the end-to-end automated system for attendance tracking and aims at optimizing Euclidean distance thresholds within deep facial embeddings. Using CNNs with ResNet-34 architecture, the proposed approach computes the 128-dimensional facial embeddings of live webcam images and compares them to a local database. At the core of this work, the evaluation of the proposed system through various Euclidean distance thresholds (0.4, 0.6, and 0.8) is performed, and the effects of those changes on True Positive and False Positive rates are analyzed. Through this process, the mathematically optimal threshold is found in terms of minimizing False Negatives without producing any false-positive outcomes. Lastly, the implementation of the developed machine learning model is showcased in the context of real-time Streamlit web-based dashboard, allowing live calculations of the system's status. (" Late" vs." On Time") and automated SMTP reporting.**

Index Terms— **Biometric tracking, convolutional neural networks, deep facial embeddings, edge computing, Euclidean distance, real-time systems.**

I. INTRODUCTION

The conventional paradigm of taking attendance within classrooms is primarily focused on manual administrative functions, which not only takes up precious teaching time but is also highly prone to human errors and potential manipulations like proxy attendance. With growing educational establishments, the requirement for automatic, seamless biometrics becomes inevitable. The facial recognition system, based on the developments in deep learning and computer vision techniques, serves as a solution to this problem. Instead of using RFID tags and fingerprint scanners, facial recognition provides an effortless way to verify students' attendance.

Still, there is a few challenges associated with implementing facial recognition systems in real-world conditions where computing capacity is limited to standard laptops, and lighting can be arbitrary. One of these important components is the classifier threshold, or the exact mathematical limit that identifies whether a newly taken picture is recognized as the same person within the dataset. This paper describes the design of a real-time attendance facial system based on Python, OpenCV, and Streamlit web interface. By artificially adjusting this threshold, we aim to quantify the trade-off between system strictness (rejecting legitimate students) and system vulnerability (accepting unregistered proxy individuals), ultimately defining the optimal operating parameters for live academic deployment.

II. RELATED WORK

Facial Recognition techniques have undergone significant improvements from localized, primitive face recognition techniques (such as Eigenface and Haar Cascade) to sophisticated, deep learning models. The use of CNNs in modern frameworks allows for the extraction of high-level feature representations. Studies on models such as FaceNet [1] and ResNet-34 [2] have proven that projections into multidimensional Euclidean spaces of facial images work well in terms of distance metrics of facial similarity. Moreover, HOG algorithms [3] have significantly boosted facial geometry extraction accuracy in noisy conditions.

Though there is ample research available analyzing such models on large-scale, controlled datasets, very little research has focused on tuning these models for low-footprint, real-time edge use cases [4]. Most research assumes that high-performance GPU clusters are readily accessible for such real-time applications. This research aims to fill this gap by performing threshold sensitivity analysis under CPU constraints, making sure that results are transferrable to institutional-level hardware.

III. SYSTEM ARCHITECTURE AND METHODOLOGY

The proposed automated attendance system comprises a pipeline of four different computer vision algorithms incorporated within a reactive user interface of the application.

A. Face Detection via HOG

The initial step in the pipeline necessitates the detection of a human face in real-time from the video feed by distinguishing the object from the rest of the scene. For improved performance in terms of computing speed, the algorithm employed for this purpose is Histograms of Oriented Gradients (HOG) [3]. HOG is capable of analyzing the distribution of edge orientations and gradient intensity, hence allowing successful extraction of a bounding box around the face irrespective of small color differences.

B. Facial Landmark Estimation

In order to compensate for the possibility of head rotation with respect to the camera, a number of regression trees is used to extract landmarks for 68 facial points that represent the edges of the eyes, nose, and mouth [5]. The estimation of these features allows for alignment and centering of the detected face prior to its being fed into the neural network model.

C. Deep Facial Embedding Generation

This aligned facial image undergoes further analysis by feeding it into a deep convolutional neural network, trained before on the ResNet-34 model [2]. The last layer of this network produces a vector of 128 dimensions called a facial embedding. This facial embedding captures all geometrical properties and structures of the face and contains compressed information for fast processing purposes and student information confidentiality.

D. Classification via Euclidean Distance

To determine identity, the system compares the live 128-dimensional vector (v_1) against the pre-computed vectors of registered students (v_2). The similarity is calculated using the Euclidean distance formula:

$$d(v_1, v_2) = \sqrt{\sum_{i=1}^{128} (v_{1i} - v_{2i})^2} \quad (1)$$

IV. EXPERIMENTAL SETUP

The main aim of the current work is to determine the optimal threshold value τ through empirical means. The experiments were carried out on an Apple Silicon architecture running at 30 frames per second. The participants involved in the tests were divided into Registered Participants (participants who had their images pre-registered in the system's internal database) and Unregistered Participants (people who could be considered potential intruders).

The tolerance parameter used by the system was varied manually within three different operation modes: Strict ($\tau = 0.4$), Normal ($\tau = 0.6$), and Loose ($\tau = 0.8$).

V. RESULTS AND ANALYSIS

The system was exposed to live video feeds across varying lighting conditions typical of a laboratory environment. Table ?? details the classification results based on the manipulated distance thresholds.

Table 1: Threshold Optimization Results

Threshold (τ)	Registered Subject	Unregistered Subject
0.4 (Strict)	FN: System routinely failed to recognize the subject.	TN: Correctly ignored the stranger.
0.6 (Default)	TP: Rapid, accurate recognition despite head tilts.	TN: Correctly ignored the stranger.
0.8(L oose)	TP: Immediate recognition.	FP: Incorrectly matched the stranger to a student profile.

[FN: False Negative, TN: True Negative,TP: True Positive, FP: False Positive]

E. Analysis of the Strict Threshold (0.4)

For τ = 0.4, however, the acceptable radius is not large enough. Any minor alterations to the surroundings, such as casting shadows due to lights placed above, alter the 128-dimensional embedding by crossing the threshold of 0.4. This results in numerous False Negatives.

F. Analysis of the Loose Threshold (0.8)

When τ is equal to 0.8, the separation curves will intersect with each other. As the tolerance increases, the number of vectors from two entirely different people that may be matched will increase. False positives will occur, which goes against the very essence of not having a proxy attended is a scalar value. A predefined threshold value τ is used to decide on the final decision. If $d \leq \tau$, a Match is declared; otherwise, it considers the input as Unknown.

G. The 0.6 Mathematical Equilibrium

This empirical evidence validates that τ = 0.6 can be considered an optimal classifier boundary. It offers sufficient mathematical flexibility to accommodate webcam errors and variations in lighting while retaining the necessary mathematical strictness to distinguish between different human identities.

VI. SYSTEM IMPLEMENTATION AND DASHBOARD ENGINEER-ING

With regard to ensuring the practicality of the most optimal τ of 0.6, the ML algorithm was applied into a stackable application on the Python language on the Streamlit framework. In this setup, the need for a complicated database system is negated due to the use of CSV manipulations. The start time of the lecture is taken as input from the user, and the model automatically inserts “LATE” or “ON TIME” tags depending on the difference in time intervals using Panda’s library. Finally, in order to produce a report on the entire process of administration, the model creates an Excel (.xlsx) file out of the daily CSV file, and sends it to the instructor via the SSL secured mail server.111

CONCLUSION

The integration of deep learning facial embeddings into classroom administration offers a highly efficient alternative to manual roll calls. However, the integrity of such a system relies entirely on the mathematical boundaries governing its classification logic. This study empirically demonstrates that when utilizing a ResNet- 34 derived 128-dimensional embedding, a Euclidean distance threshold of 0.6 provides the optimal equilibrium between user convenience (minimizing False Negatives) and institutional security (eliminating False Positives). By embedding this optimized pipeline within a Streamlit architecture, this project proves that high- accuracy biometric tracking can be successfully deployed on lightweight edge hardware without relying on expensive cloud infrastructures.

ACKNOWLEDGMENT

The author thanks the Department of Computer Engineering at Bharati Vidyapeeth College of Engineering, Pune, for providing the academic infrastructure and support for this research.

REFERENCES

[1] F. Schroff, D. Kalenichenko, and J. Philbin, “FaceNet: A unified embedding for face recognition and clustering,” in *Proc. IEEE Conf. Computer Vision and Pattern Recognition (CVPR)*, 2015, pp. 815–823.

[2] K. He, X. Zhang, S. Ren, and J. Sun, “Deep residual learning for image recognition,” in *Proc. IEEE Conf. Computer Vision and Pattern Recognition (CVPR)*, 2016, pp. 770–778.

[3] N. Dalal and B. Triggs, “Histograms of oriented gradients for human detection,” in *Proc. IEEE Computer Society Conf. Computer Vision and Pattern Recognition (CVPR)*, 2005, pp. 886–893.

[4] S. A. Gite and A. Agrawal, “Automated attendance system using face recognition,” in *Proc. Int. Conf. Computational Intelligence and Computing Research (ICCIC)*, 2020, pp. 1–5.

[5] D. E. King, “Dlib-ml: A machine learning toolkit,” *Journal of Machine Learning Research*, vol. 10, pp. 1755–1758, 2009.